UNIVERSITY OF MICHIGAN

# Tell don't just show: Narratives improve insight more than interactivity in communicative visualizations

by
**Abhraneel Sarma**

A thesis submitted in partial fulfillment for the degree of Master of Science in Information

Completed under the supervision of
Matthew Kay
School of Information

April 2018

## Declaration of Authorship

I, ABHRANEEL SARMA, declare that this thesis titled, "Storytelling in Information Visualization: Does higher engagement lead to higher recall?" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ______________________________

Date: ______________________________

# Abstract

Communicative visualizations incorporating narratives and/or interactivity are increasingly commonplace. However, the relative importance of narrative versus interactivity in improving readers' understanding is unclear. We designed visualizations which vary in presence of interactivity (static or interactive) and narrative (non-narrative or narrative), presented them to Turkers, and measured recall using an 11 item True/False questionnaire. We find a weak positive effect on recall—an increase of ~8.4 percentage points (95% CI: [4.5, 12.7])—from the presence of narratives, but little or no effect from interactivity (95%CI: [-1.1, 5.1]). We argue that narratives can better facilitate insight generation for viewers of communicative visualizations, but that interactivity may not. We discuss implications for the broader definition of information visualization itself: contrary to well-known existing definitions, we question whether interactivity is a necessary component of any information visualization, rather than a subordinate component that is useful if and when a design calls for it.

## Acknowledgements

I would like to take this opportunity to thank the people who have contributed to the successful completion of the Thesis. My thesis project was highly collaborative and I am grateful to the people around me who have helped me by advising me and supporting my research, which allowed me to successfully complete this project.

I would like to acknowledge the support of my advisor, Matthew Kay, who has been my mentor for the past 2 years and whose constant guidance, support and encouragement has made this thesis possible. Matt offered me the initial opportunity to be a part of a research project in my first semester at UMSI, and has helped me in developing the skills to conduct research with good practices. Throughout my thesis, Matt has always offered unique perspective which made the experience very insightful, and whenever necessary, pushed me to strive harder and look at the bigger picture.

I would also like to thank several other members of the faculty at UMSI: Chris Brooks, David Jurgens and Chris Teplovs, who have been available to lend their expertise whenever I needed advice.

I am also grateful to my friends and fellow students at UMSI. Raymond Su, Raden Tonev, Deepak Krishnan, Tait Chamberlain, Mauli Pandey, Paul Hur, Jason Lam have all always engaged me in some thoughtful discussion whenever I felt stuck at any point in my project. Gary Cheung bailed me out by readily sharing his expertise and code for one of the visualizations, when I was really struggling.

Finally, I would like to thank my parents who always have believed in me and supported me in all my decisions. Without them, I do not think any of this would have been possible.

# Contents

# 1 Introduction

# 2 Introduction

The past several years has seen the growth of author-driven, narrative-based visualizations, particularly in data journalism. The primary goal of such visualizations is the communication of an intended message to the audience. By presenting insights from the data interwoven with a story [7], these visualizations can make it easier for readers to make sense of the data, thereby making the data more accessible. Communicative visualizations use narratives and interactivity as two design strategies, to facilitate this insight generation process.

Interactivity is considered an essential component of visualization, including communicative visualizations. It allows the user to perform operations on the data and extract meaning. This is evident from Card, Mackinlay, and Schneiderman's oft-quoted definition of information visualization: "*the use of computer supported, interactive, visual representations of abstract data to amplify cognition*" [12]. However, this perspective developed with the goal of supporting exploratory data analysis. According to Aisch [1], a large percentage of the readers of communicative visualizations do not make use of interactive elements in those visualizations. Aisch's remarks sparked a debate about the importance of interactivity in such visualizations [2, 6]. But the effects of interactivity in communicative visualization effect on insight have not been systematically studied.

At the same time, some authors [16, 34] have discussed the potential value of narratives in communicative visualization. They claim that using narratives and storytelling in visualizations facilitates the efficient communication of large amounts of information. This, they theorize, would make the process of gathering insights about the data by the viewer more effective.

It is unclear how, if at all, these two design strategies—narrative and interactivity—impact the amount of insight that users gain from a communicative visualization. Thus, the primary contribution of this work is **to study the effect of the use of narrative and interactivity, as two different visualization design strategies, on insight gained by the reader.**

Measuring insight to evaluate visualizations poses an additional challenge. Insight has not been consistently defined in the visualization literature. Some recent work which evaluated communicative visualizations has used metrics like engagement [4, 10, 17, 35]. Boy et al. [10] investigated whether adding an introductory narrative component—a short story about the visualization—increased user engagement. They measured engagement through total time spent on, and the number of meaningful interactions with, the visualization. They found that the addition of an *introductory narrative component* to a visualization did not increase user engagement, and concluded that the narrative component does not motivate users to engage in data exploration, which might then lead to less insight gained.

But engagement is an indirect measure of insight. It does not necessarily indicate whether the user was able to gather the intended message from the visualization Instead, we propose to measure insight directly by quantifying factual and conceptual knowledge gained from the visualization [3]. We use a questionnaire which evaluates participants recall and comprehension of the data. This gives us a more direct measure of insight, which allows us to better evaluate the effectiveness of a communicative visualization. Mahyar et al. [25] pose a similar argument, saying that measuring insight is a challenge, and simple engagement metrics like interactions may be insufficient for measuring user engagement as they are only indicative of low-level engagement. They propose measuring user performance on higher level cognitive tasks involving recall, comprehension, analysis and synthesis of the data.

In this paper, we present the results of an online study conducted on Amazon's Mechanical Turk with 389 participants through which we evaluate the effects of different visualization strategies—interactivity (static or interactive) and narrative (present or absent)—on recall and comprehension. In a pre-registered, confirmatory analysis, we measured recall and comprehension by using the probability of correctly answering a question on an 11 item True / False questionnaire based on the visualization. We find that:

- The presence of a narrative resulted in higher recall of visualization content. A participant in a narrative condition was, on average, 8.5 percentage points (95% CI: [4.5, 12.7]) more likely to answer a question correctly, compared to a participant in a non-narrative condition. This effect is equivalent to answering, on average, one more question correctly on the 11 item questionnaire.

- The presence of a interactivity likely does not have a large effect on insight. The mean effect was an increase of 1.9 percentage points

(95% CI: [-1.1, 5.1]) on the probability of answering a question correctly.

We also conducted some followup exploratory analyses, including an analysis of the relationship between engagement and insight. We found that spending more time on a visualization may result in more insight; however, performing more interactions with the visualization may not result in more insight. This indicates that interactions (such as clicks) may not not be a good metric to measure insight—which, ultimately, is our outcome of interest.

Our results suggest that, for communicative visualizations, using a narrative component may be more important for helping viewers gain insight than interactivity. Thus, we question whether interactivity should be considered an essential component of communicative visualizations, rather than a component to be employed as and when needed.

# 3 Background

In order to systematically test the differences in the impact of narrative and interactivity on insight, we need easy-to-operationalize definitions of narrative and interactivity. Therefore, we look at how *narrative* and *interactivity* in visualization have been defined in prior literature, which we will use to design visualizations which clearly have or do not have each property. In order to be able to measure *insight* directly, we need to understand how insight has been defined in the context of InfoVis and identify methods to measure insight directly. Finally, we review prior work which study the effects of visualization on metrics such as engagement, memorability, recall.

## 3.1 Narratives in Visualization

Narrative visualizations have become widespread in the past decade, particularly in journalism. Segel and Heer first defined a design space for narrative visualizations and placed them along a spectrum between reader-driven and author-driven approaches [33]. They identify three commonly used schemas on this spectrum, of which we focus on the first, *the Martini Glass structure*. The Martini Glass structure is a primarily author-driven approach where the author builds a narrative, in a linear order, around the visualization using questions, observations and text passages. At the end of the narrative, the viewer can freely interact with and explore the visualization. Thus through the author-driven narrative, the designer tries to communicate the primary insights about the data to the viewer.

Following Segel and Heer's analysis of the design space of narrative visualization, there have been several attempts at defining this genre of visualizations. Kosara and Mackinlay, defined a story or narrative as *"an ordered sequence of steps, with a clearly defined path through it"*. When the steps primarily consist of information visualizations, they are termed as narrative visualizations [22]. Hullman and Diakopoulos define narrative visualization as *a style of visualizations which combine persuasive, rhetorical techniques for explanation with interaction techniques for exploration by the user*. [19]

Lee et al., [23] attempt to consolidate what constitutes a *visual data story* by defining them as:

- *They consist of a set of story pieces, i.e. facts backed up by data.*
- *Most of the story pieces are visualized to support one or more of the intended messages. The visualization includes annotation or narration to clearly highlight and emphasize this message, and to avoid ambiguity.*
- *Story pieces are presented with a meaningful order or connection between them to support the author's higher level communication goal.*

**In this paper**, we use Lee et al.'s definition to classify a visualization as a narrative visualization. Although Lee et al. uses the term *visual data story*, it is clear that they are referring to the same genre of visualizations as Segel and Heer's *narrative visualization* [33], and we use the latter term to refer to this genre of visualizations in the paper. We use this definition because it encompasses the spectrum of author-driven and reader-driven narratives in visualization, where the ordered sequence of steps is essential to build the narrative.

### 3.2 Interactivity in Visualization

Although interactivity is commonly used in information visualizations, there have been continued efforts to precisely define *interactivity* in the context of InfoVis. A lot of these definitions are derived from how interaction techniques are defined for human–computer interaction (HCI):

> *The interaction component involves the dialog between the user and the system as the user explores the data set to uncover insights. The interaction component's roots lie in the area of HCI.* [37]

Yi et al. go on to define the following seven categories as interactions for InfoVis: *select*, *explore*, *reconfigure*, *encode*, *abstract/elaborate*, *filter*, and *connect*. Ziemkiewicz and Kosara, make a distinction between trivial and non-trivial interaction, where non-trivial interaction facilitates *active readability* (allowing the user to actively seek information) by allowing the reader to make changes to the parameters of the visual mapping itself [38].

**In this paper**, using the definition of non-trivial interactivity, we consider interactivity in visualizations to consist of only the following five categories identified by Yi et al.— *reconfigure, encode, abstract/elaborate, filter*. We consider *select* and *explore* as trivial interactions. We make this distinction because of the following reasons: (1) *select* includes interactions such as hover and other tooltip interactions, which don't change the visual mapping in any way and are often used to give precise information about a data point, and may not contain any additional information. (2) Interactions such as scrolling or panning, (classified under *explore*) are often used to overcome the constraints of a screen and hence should be considered trivial interaction. For e.g., the Earth temperature timeline [28] visualizes the temperature for the last 20,000 years is a single, very tall, static image. Scrolling through this visualization would, by definition, constitute performing an *explore* interaction; but this interaction—scrolling through a static image to overcome the constraints of the browser—is not changing the mapping of the data in anyway, and such interactions should be considered trivial.

## 3.3 Visualization Insight

It has been argued that the main goal of visualization is to provide insight [12, 15, 30]. However, the use of the term *insight* in visualization literature has not always been consistent. Saraiya et al. [32], define insight as*"an individual observation about the data by the participant—a unit of discovery"*. North [30] acknowledges that defining insight is very challenging, and describes characteristics of insights—complex, deep, qualitative, unexpected, relevant. Yi et al. [36], describe how people gain insights when using a visual analytic tool, and identify different processes which people use while exploring the data to gather insights.

Chang et al., argue that the way insight has been described in prior literature in visualization is more or less similar to *"units of knowledge"*, which is different from how insight is understood in cognitive science research [14]. In this paper, consistent with how insight has been interpreted in visualization literature, we treat insight as knowledge and information that can be gained (from a visualization).

## 3.4 Measuring insight from Visualization

Recently, several studies have looked into how different types of visualizations affect the reader, in an attempt to identify low-level features which make a visualization more engaging, memorable or easily recognizable. Bateman et al. found that the use of chart junk (or embellishments) in visualizations do not negatively affect participants' accuracy of interpretation, and can even help them perform better on long-term recall tasks [4]. Consistent with this result, Borkin et al. found that features such as human-recognizable objects and the use of more colors can make a visualization more memorable, even when participants were exposed to the visualization for just one second [9]. In a subsequent study, Borkin et al. found these results to be consistent even on prolonged exposure to the visualization (10 seconds) [8]. They also find that people recall more details about visualizations which are more recognizable. Haroz et al., in their evaluation of the effects of pictographs in visualizations found that these were recalled more accurately during demanding tasks, and also led to users paying more attention to the visualization [17]. All of these studies looked at visualizations as static representations of data, and identified elements within the visualizations which make them easily recognizable and memorable. However, these studies did not look at the effect of the presence of design strategies such as narratives or interactivity.

Other studies have looked into factors which affect viewer engagement with a visualization. McKenna et al., identify characteristics of visual narrative flows in data-driven stories, and study the effect of two commonly used flows—stepper vs scroller-driven flows—on readers. They used a self-reported score using a validated questionnaire to measure user-engagement. The authors did not find any difference on engagement between the use of steppers and scrollers [26].

In a study which is perhaps the most relevant to this work, Boy et al., performed a large field study where they measured engagement, using the total time spent on the visualization and total number of (hover and click) interactions, to study if the inclusion of an introductory narrative component motivates viewers to engage and explore the data to a greater extent [10]. In their study, users were placed in one of two conditions: a storytelling version which had an introductory narrative about the data presented using a stepper-flow, followed by an exploration section which allowed the users to interact with the visualization; and a non-storytelling version which directly presented users with the explore section of the interactive visualization. They find that storytelling did not lead to greater engagement with the visualization, based on these metrics. Since the presence of an introductory narrative component makes readers less inclined to explore the data, the authors conclude that readers might be gathering fewer insights from the data.

Most of the studies discussed above use engagement to evaluate a visualization. Although higher engagement, and exploration, *may* result in users gaining more insight from the data, this is an indirect measure. Thus, it remains unclear how these design strategies affect insight gained by the viewer. We propose instead to measure insight more directly. Since insight in the context of InfoVis is interpreted as *"units of knowledge"*, one way of measuring insight could be to evaluate the knowledge gained by the viewer using metrics which represent learning. According to the revised Bloom's taxonomy, recall/recognition and comprehension/understanding are two levels of the cognitive domain that can be used to test knowledge at the factual and conceptual level [3]. Hence, we use recall and comprehension as metrics to measure the effects of narratives and interactivity on insight.

# 4 Design

## 4.1 Design of Visualizations

To systematically test the effects of interactivity and narrative on the user, we designed visualizations which either have or do not have narrative and/or interactivity. We use the definitions of narrative and interactivity to propose a design space for classifying information visualizations (Figure 1), with respect to narrative and interactivity, where each visualization can vary in degree of narrative (how much the author guides the viewer through the insights that can be gathered from a visualization) and interactivity (how many, and to what extent, of the categories of interactivity it supports). By definition, all visualizations would fall into one of the four quadrants in this design space.

We then created a catalog of visualizations by going through websites such as Visualizing.org[1], which maintains a catalog of professionally produced visualizations as well as websites of news agencies such as The New York Times, The Washington Post, Bloomberg etc., which maintains a list of visualizations that were published by their respective graphics department. We restricted ourselves to reviewing and selecting only professionally produced visualizations to ensure high quality. We identified 40 visualizations, and mapped each onto the defined design space. From

[1] https://www.visualizing.org/

this catalog, we selected four visualizations using the following considerations: (1) all of the visualizations possessed a strong narrative component; (2) we felt the visualization would be suitable for adaptation to the other quadrants of the design space; (3) all the visualizations depicted data that was readily and publicly available so that they can easily be reproduced. We selected four visualizations to ensure that the any observed effects would be consistent across visualizations and not due to the properties of a specific visualization.

Each of the four visualization were redesigned to have one version in each condition: *no-Nar+no-Int* (baseline condition), *Nar+no-Int*, *no-Nar+Int* and *Nar+Int*, resulting in 16 unique combinations of visualization and design strategy. Since we evaluate the visualizations using recall and comprehension, we needed to ensure, across each version of a visualization, consistent encoding and layout, and equal expressiveness—the visual encoding of the data expressed equal amount of information [29].

To achieve this, we followed a structured and iterative design process to reproduce each visualization, which is described in Figure 1. We first followed this approach to create paper-based mock-ups for each version of the visualization. The *Nar+Int* versions of each adapted visualizations used a Martini-glass structure [33], and consisted of a nar-

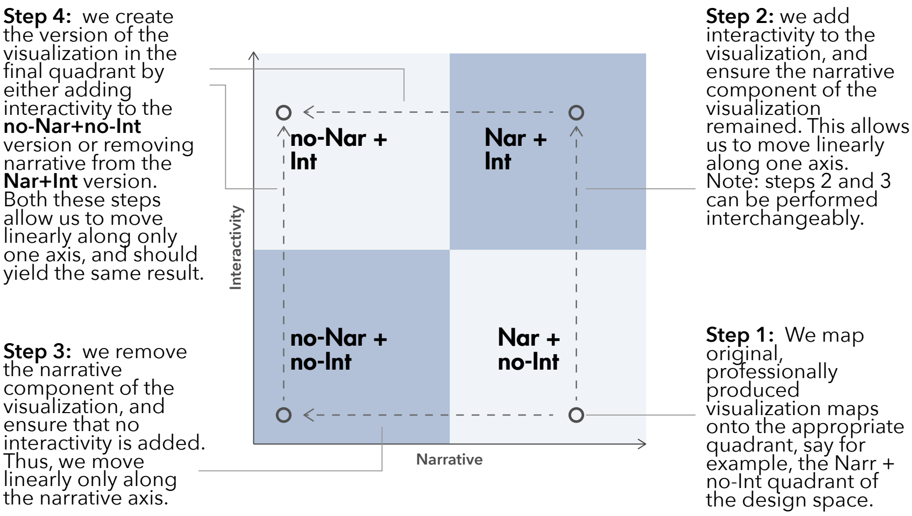


**Figure 1: Design Space for classifying visualization along a spectrum for narrative and interactivity; we use this design space and a systematic approach described in the figure to adapt a visualization from its original mapping onto the design space to the other quadrants**

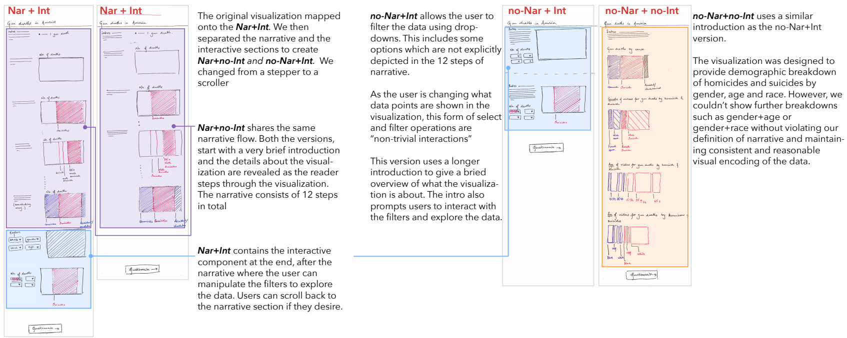


**Figure 2: Adapting a visualization from its original mapping onto the design space to other corners of the design space. Here, we show the adaptation process for *Gun deaths in America* visualization. The original version mapped onto Nar+Int quadrant. We iteratively designed versions which mapped on to other quadrants using the definitions of narrative and interactivity. We developed using HTML, JavaScript and D3.js**

rative section, followed by an interactive section. The narrative section allowed the user has to scroll or step through through. In the interactive section, the user can perform non-trivial interactions with the visualization. The *Nar+no-Int* versions consisted of just the narrative part of the corresponding *Nar+Int* version. Similarly, the *no-Nar+Int* versions consisted only of the interactive section of the corresponding *Nar+Int* version. Finally, the *no-Nar+no-Int* version of the visualization consisted of non-interactive visualizations, and the visualizations were presented in a logical order (2). However, we ensured that this order did not implicitly imply a narrative to the viewer.

In what follows we will briefly describe the 4 visualizations. The visualizations can be accessed here: `https://visinsights.github.io/`. The contents of the website have been anonymised and we do not track IPs or any visitor information.

## 4.2 The Visualizations

*Gun Deaths in America*[2] [13] was a part of fivethirtyeight's project which explored the 30,000+ annual deaths in the US which are caused by gun violence. The original graphic mapped onto *Nar+Int* quadrant of our design space. Stepper buttons allowed the reader to advance through the linear narrative. It plotted the data for the year 2014 as a dot matrix plot

[2]https://fivethirtyeight.com/features/gun-deaths/

with each dot representing a single victim. It encoded other information such as type of death (Suicide, Homicide, Accidental or Undetermined), and demographic information such as race, gender and age. In the final scene of the sequence, users could explore the data using a set of filters.

*Bloomberg Carbon Clock*[3] [31] is a real-time estimate of the global monthly atmospheric $CO_2$ level. This is accompanied by a narrative visualization. The visualization consists of a time-series line chart, which contrasts the rate of increase of atmospheric $CO_2$ levels over the past 60 years to historical trends over the past 12,000 and 800,000 years. The original visualization used a continuous scroll to allow the reader to advance through the narrative. At the end of the narrative, users could toggle through the views (past 3 years, 60 years, 12,000 years or 800,000 years) using buttons.

*Spread of measles*[4] [18] is a animated dot-matrix style visualization which depicts how 10 hypothetical communities with different vaccination rates will be affected when they come into contact with a person infected with measles. To provide a real-world link, they picked some US cities which have comparable vaccination rates to the plots shown. The original visualization mapped on to the *no-Nar+no-Int* quadrant and is accompanied by two passages. One describes the graphic and the assumptions made. The other is a journalism piece on the 2015 outbreak of measles in the US. We created a narrative using the passages in the article. We used the county-level US kindergarten MMR vaccination rate assessment data [21] to create the interactive sections of the *Nar+Int* and *no-Nar+Int* design which allowed users to change the vaccination rate using a slider and run the simulation.

*US spending on healthcare*[5] [27] is a two axis parallel coordinate interactive visualization which compares the cost of healthcare along 7 different metrics to the corresponding quality of care measured by 8 metrics in 35 OECD countries. It mapped onto the *no-Nar+Int* quadrant of our design space. Users can use two drop-down menus to change the metrics displayed along the two axes. An introductory passage precedes the visualization and introduces the reader to the data. It contrasts the high-spending on healthcare in US and its relatively low scores on most quality metrics. We created a linear narrative from the introductory passage

[3] https://www.bloomberg.com/graphics/carbon-clock/

[4] https://www.theguardian.com/society/ng-interactive/2015/feb/05/-sp-watch-how-measles-outbreak-spreads-when-kids-get-vaccinated

[5] https://www.bloomberg.com/graphics/2017-health-care-spending/

for the *Nar+Int* and *Nar+no-Int* versions.

For each visualization, we ensured that the different versions were as consistent as possible. We used the same visual encoding and visualized the same data attributes across each version. We then developed the 4 versions for each visualizations as standalone web-pages using D3.js. To ensure that each of these visualizations had the same degree of expressiveness, we conducted an *informal pre-test pilot study* with 6 graduate students of HCI who had completed one graduate-level course on InfoVis at a large public university. We presented each student with two versions of a visualization; and asked them to compare the expressiveness between the two versions. We used open-ended questions to prompt exploration so that they were able to identify all the data attributes encoded. The goal of this pilot was to ensure that each visualization encoded the same amount of information a viewer can gain—as insight—and which can be evaluated using a questionnaire. We made further changes to the design, based on their direct feedback.

# 5 Method

## 5.1 Research Questions

Through this study we attempted to answer the following, pre-registered[6] primary research questions:

1 (a) What is the effect of *interactivity* on recall and comprehension?

1 (b) What is the effect of *narrative* on recall and comprehension?

For both the research questions, the probability of an average participant answering an average question correctly is used as the dependent measure. The predictors are the two variables, *interactivity* (present or absent) and *narrative* (present or absent), along with their interaction term.

In addition, we also attempted to answer the following exploratory questions:

2 (a) What is the effect of the presence of a *narrative* on the total time spent on the visualization?

2 (b) What is the effect of the presence of *interactivity* on the total time spent on the visualization?

The dependent variable is the duration of total time spent on a visualization; the independent variables are the same.

3 (a) What is the effect of total time spent on the visualization on recall and comprehension?

3 (b) What is the effect of the number of meaningful interactions (clicks) on recall and comprehension?

Here, the dependent measure is the number of correctly answered questions by each participant and the independent variables are time spent and number of clicks respectively.

[6]The anonymous pre-registration document can be found in the supplementary materials

## 5.2 Measuring Insight

To answer our confirmatory research questions we created an 11 item True / False questionnaire for each visualization[7], which aimed to test recall and comprehension of the participants. Based on our learning goals, we used the revised Bloom's taxonomy [3] to devise a set of questions which can be used to measure how much participants remembered (this includes recall, recognition of facts and basic concepts) and understood (this includes interpretation, inference, comparison of ideas or concepts) the information presented through a visualization. We ensured that all the questions could be answered by participants in any condition.

One question on the questionnaire was designed as an attention check question to which the answer was clearly false if the reader understood the topic of the visualization (eg., the attention question for the *Gun deaths in America* visualization was "None of the deaths shown in the graphic were caused by an incident which involved a gun"). The other 10 questions were based on the data presented in the visualization and we ensured, by making the visualizations equally expressive, that they can be inferred in each of the four conditions. In addition, we decided to reject participants who answered *True* to all the questions, or *False* to all the questions.

We ran another pilot study using the final experimental design and recruited participants via email lists from a large public university. The pilot helped identify poorly worded sentences, a rough estimate of the average number of questions that we can expect a participant to correctly answer in the final study in order to determine pay.

## 5.3 Other metrics

To answer our exploratory research questions, 2 (a), 2 (b), and 3 (a), 3 (b), we also collected data on the amount of time spent by the participants on the webpage for all the four conditions. In the two conditions with interactivity (*no-Nar+Int* design, *Nar+Int*), we measured the number of non-trivial interactions with the data. Non-trivial interactions were measured by counting the number of clicks used to filter the data (Gun Deaths in America), change the view (Carbon Clock), change the vaccination rates using an input slider (Measles), change the dimensions (US healthcare).

[7]The questionnaires can be found in the supplementary materials

## 5.4 Study Design and Procedure

In our study, the fixed (population-level) effects are the presence (or absence) of *narrative* and the presence (or absence) of *interactivity* and their interaction. This results in four conditions. In addition, we have random (group-level) effects for *visualization*.

We use a completely between subjects design where each participant was placed in one condition. Each participant was presented with one visualization and then answered an 11 item questionnaire.We performed a power analysis to determine the number of participants to recruit for the final study. Based on this, we decided to recruit 100 participants per condition. We pre-registered our Bayesian regression model, along with how we were planning on defining outliers (rejection criteria as mentioned above) using AsPredicted.org before collecting and analyzing our final dataset. We launched the study as a single HIT on Amazon's Mechanical Turk. Participants were instructed to go through each step of the visualization carefully and then proceed to the questionnaire. Participants were not allowed to go back to the visualization once they reached the questionnaire, and were informed of this in the beginning of the study.

We recruited participants on MTurk with the qualifications that they have a prior HIT approval rating of 98% and have completed at least 500 prior HITs. Each participant was given a base pay of \$0.75 and informed that they will receive a bonus of \$0.2 for every question that they answered correctly. We introduced the incentive for answering questions correctly to motivate participants to spend time on the visualization, with the goal to simulate the intrinsic motivation that a user on the web might have to visit a data visualization in a news website.

The average time for the workers to finish the HIT was slightly under 8 mins and the average payoff was \$2.30. In total we received 389 responses: 97 (83)[8]in static, non-narrative condition; 100 (86)[8] in interactive only condition; 105 (95)[8] in the narrative only condition; 87 (80)[8] in interactive and narrative condition. We rejected two participants outright as they answered *True* to all the questions. The other 43 participants failed just the attention check question. We noticed that most of the participants who failed just the attention check question saw the Carbon Clock visualization (31). We suspect that the attention check ques-

[8]the numbers within parenthesis indicate the number of responses after filtering out outliers based on our pre-registered criteria.

tion for this visualization may have been more difficult than the others. Hence, we performed our primary analysis twice—first excluding the participants according to our pre-registered model; we then re-ran our model including the 43 participants who were excluded for failing the attention check question. The results for both the analyses were similar. In this paper, we report our pre-registered analysis. The results for the analysis including the data of the participants who failed the attention check question can be found in the supplementary materials.

## 5.5 Model

After Kay et al. [20], we implemented a Bayesian multilevel logistic regression model for our confirmatory analysis using the brms package in R [11]. Our model can be represented using the lmer formula syntax [5] as follows. Please refer to the supplementary materials for the complete model.

$$
\begin{aligned}
correct \sim\ & interactivity \times narrative + (1|participant) \\
& + (interactivity \times narrative|visualization) \\
& + (1|question)
\end{aligned}
$$

Our model calculates the probability of an average participant answering an average question correctly, using *interactivity* and *narrative* and their interaction as population-level effects. We include group-level effects for *participants*, *questions* and *visualization*. We model each *participant* using a varying intercept because participants will have differing abilities resulting in different baseline probabilities of answering a question correctly. We model each *question* using varying intercepts, as they may be of different difficulty. Finally, the effect of the variables—*narrative*, *interactivity* and their interaction term—may also be different based on the *visualization*, which we model using varying slopes and intercepts.

Mathematically, our model can be written as:

$$y_i \sim Binomial(n_i, p_i)$$

$$
\begin{aligned}
logit[p_i] = &\alpha + \alpha_i + \alpha_{vis[i]} + \alpha_{j,vis[i]} + \\
&(\beta_{int} + \gamma_{int,vis[i]})x_{int} + \\
&(\beta_{narr} + \gamma_{narr,vis[i]})x_{narr} + \\
&(\beta_{narr \times int} + \gamma_{narr \times int,vis[i]})x_{narr}x_{int}
\end{aligned}
$$

$$\alpha_i, \alpha_{j,vis[i]} \sim \mathrm{N}(0, \theta)$$

$$\theta \sim \mathrm{t}(3, 0, 10)$$

$$\alpha, \beta_{int}, \beta_{narr}, \beta_{narr \times int} \sim \mathrm{N}(0, 1)$$

$$\begin{bmatrix} \alpha_{vis[i]} \\ \gamma_{int,vis[i]} \\ \gamma_{narr,vis[i]} \\ \gamma_{int \times narr,vis[i]} \end{bmatrix} \sim \mathrm{MVN}(0, \Sigma)$$

$$\Sigma \sim \begin{pmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_{int} & 0 & 0 \\ 0 & 0 & \sigma_{narr} & 0 \\ 0 & 0 & 0 & \sigma_{int \times narr} \end{pmatrix} R \begin{pmatrix} \sigma_\alpha & 0 & 0 & 0 \\ 0 & \sigma_{int} & 0 & 0 \\ 0 & 0 & \sigma_{narr} & 0 \\ 0 & 0 & 0 & \sigma_{int \times narr} \end{pmatrix}$$

$$\begin{pmatrix} \sigma_\alpha & \sigma_{int} & \sigma_{narr} & \sigma_{int \times narr} \end{pmatrix} \sim \mathrm{t}(3, 0, 10)$$

$$\mathrm{R} \sim LKJcorr(1)$$

Variables:

- $i = 1 \ldots I$ indexes persons (respondents of the survey).

- $j = 1 \ldots J$ indexes items / number of questions on the questionnaire. Here, $J = 11$ .
- $vis[i] = 1 \ldots K$ indexes the number of visualizations for a participant, i. Here $K = 4$.
- $y_i \in 0, 1$ is the response by participant i.

Parameters:

- $\gamma_{vis[i]}$ is the coefficient for random slope for a visualization, $m$.
- $a_i$ is the random intercept due to the ability of a person $i$.
- $a_{vis[i]}$ is the random intercept due to each visualization.
- $a_j$ is the random intercept due to the difficulty of each question $j$, in each visualization.

# 6 Results

## 6.1 Confirmatory Analysis

Figure 3 shows the posterior density, the mean point estimate, and 66% and 95% quantile credible intervals of answering a question correctly in each of the four conditions, and the mean difference of the effects of *narratives* and *interactivity* respectively on the probability of getting a question correct by the viewer. We find that the presence of a narrative has a weak but positive effect on recall and comprehension, increasing the probability of answering a question correctly on average by 8.5 percentage points (95% CI: [4.5, 12.7]). The mean effect size is of the order of getting one more question correct, for an average participant, for a particular visualization. On the other hand, presence of interactivity likely has little or no effect—interactivity increased the probability of correctly answering a question by 2 percentage points (95% CI: [-1.1, 5.1]).

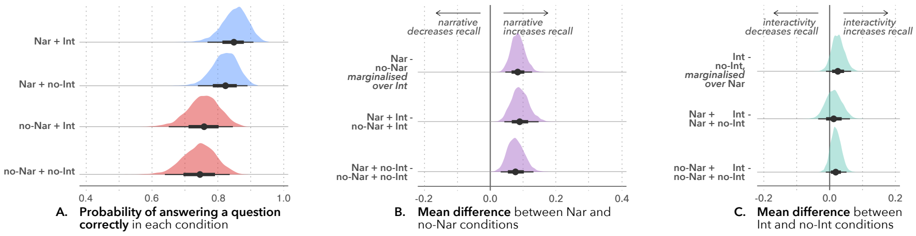


**Figure 3: Posterior density, mean point estimate, 66% and 95% quantile credible intervals for the probability of correctly answering a question, marginalised over the group-level effects of visualization. A. shows the estimates for each condition; B. and C. show the mean differences for the effect of narrative and interactivity respectively. The results shown excludes participants who failed the attention check (pre-registered criteria).**

Figure 4 shows the mean difference for the two variables conditions, in the probability of answering a question correctly, separately for each visualization (taking into account the group level variances). We can see that the presence of narrative has a consistent effect across visualizations. The effect of interactivity is also consistent and does not seem to indicate a positive effect on answering a question correctly.

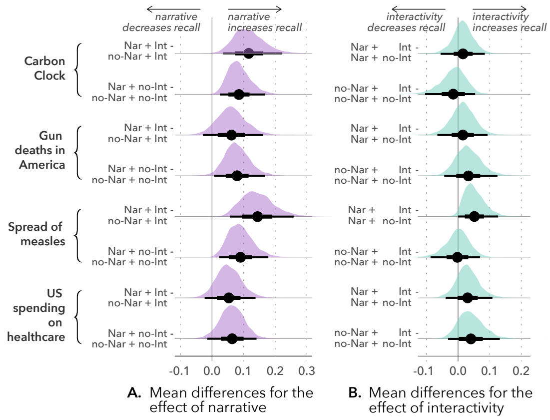


**Figure 4: Posterior probability estimates for the mean differences between the conditions, showing the group-level effects for each visualization**

## 6.2 Exploratory Analysis

To answer our exploratory questions 2 (a) and 2 (b), we fit a Bayesian multilevel model using the total uptime as the dependent variable and *narrative* and *interactivity* as the predictors. Figure 5 depicts the results, which shows that narrative may have a small positive effect on uptime (Mean: 55s, 95% CI: [27s, 82s]). On the other hand, we find that there's likely little or no effect on uptime due to the presence of interactivity (Mean: 18s, 95% CI: [-10s, 46s]). However, we should note here that there is a lot of uncertainty in our estimates, as evidenced by the wide 95% intervals (the width of the intervals are ~55s), and thus these results should be considered cautiously.

$$\begin{aligned} uptime \sim & interactivity \times narrative \\ & + (interactivity \times narrative | vis) \end{aligned}$$

To answer our exploratory questions, 3 (a) and 3 (b), we fit two linear models. We use the number of questions correctly answered as the dependent variable for both the variables. We use the log of the duration of time spent on the visualization as the independent variable for the first model; and the number of non-trivial interactions with the visualization as our independent variable for the second model. Figure 6A indicates that there may be a slight positive correlation of time spent with the number of questions correctly answered by our participants. However, in Figure 6B we see that there may not be any positive correlation

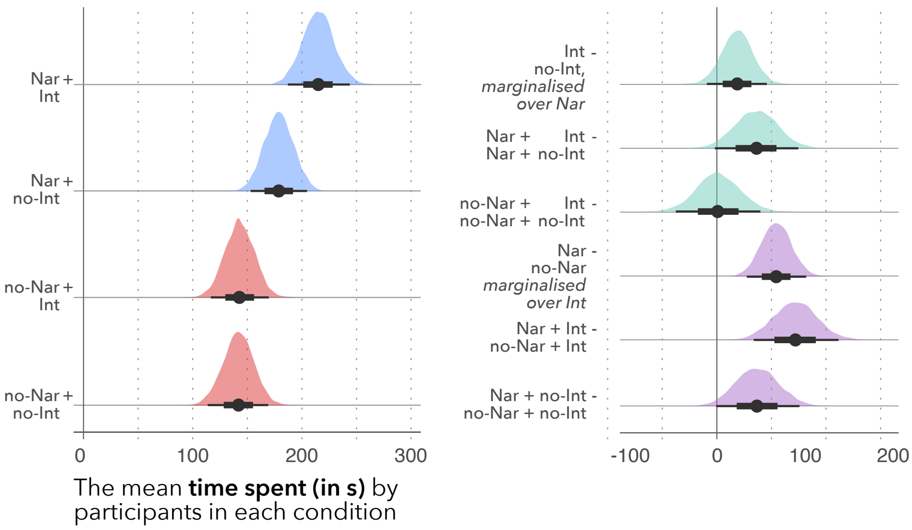


**Figure 5: Posterior probability estimates for the time spent by an average participant in each condition, and the mean differences between the conditions**

between the number of non-trivial interactions and the number of correctly answered questions.

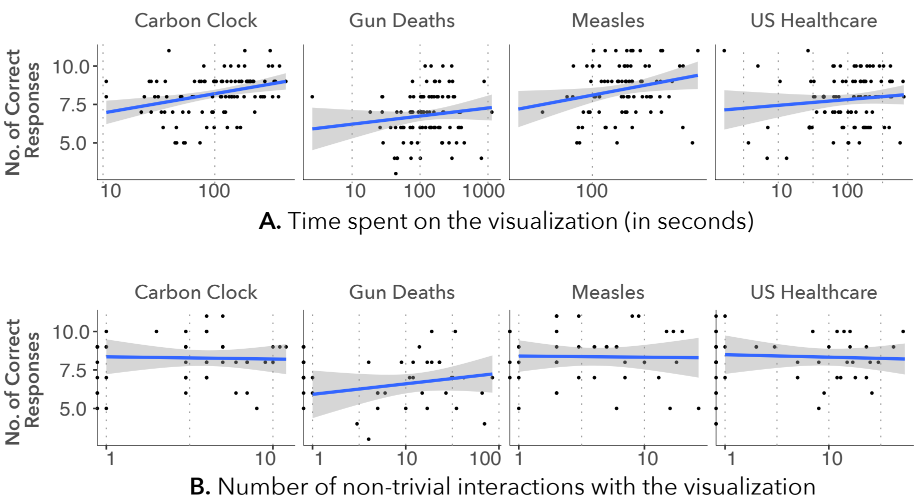


**Figure 6: The effect of (A.) the total time spent in the visualization and the (B.) the number of interactions performed by a participant on the number of questions answered correctly**

# 7 Discussion

Measuring the effectiveness of different visualization design strategies on the insight can be challenging. Boy et al. [10] argued that insight is generated through data exploration, and the tendency to explore the data can be increased by increasing user engagement. To measure engagement, they used two metrics—time spent on the visualization and the number of semantic operations. However, it is not clear that higher engagement will necessary lead to greater insight. Spending time on, and performing interactions with, a visualization does not necessarily mean that the user has assimilated the information that the designer is trying to convey to them, and thus, may not be correlated with insight. Instead, if a visualization was effective in allowing the readers to generate insights, they should understand the data better (and may even be able to do so more quickly). Hence, recall and comprehension would be more direct measures of insight.

## 7.1 The effect of interactivity and narrative on insight

North highlights the need for "*newer evaluation methods which aim to measure insight directly*" [30]. We designed our experiment to measure users' ability to recall, and their comprehension of, information from the visualization by calculating the probability of answering a question correctly. We find that narratives may be able to better communicate the insights from the data to the audience (Figure 3). This could be because a narrative places greater emphasis on the main insights that can be gathered from the data. This reduces the barrier to gather insights, since the viewer does not need to engage in data exploration to gather relevant insights. By contrast, in our task, interactivity showed little effect on insight.

Yet, interactivity is widely considered to be an essential part of information visualization. In his *Information Visualization Manifesto*, Manuel Lima, citing Card et al. [12], argues that interactivity is key for information visualization [24]. Yet making any particular type of feature essential to a visualization contradicts the first point in the manifesto: *form follows function*. From a user-centered design perspective, interactivity (*form*) should only be adopted if user needs (*function*) dictate it. When the goal of a visualization is to communicate specific insights to the user, interactivity should not be considered a necessary component unless it furthers that goal.

This should not be interpreted as interactivity having no benefit, necessarily. As Gregor Aisch describes [2], even though a large percentage of the consumers of communicative visualizations do not use interactive features, there may be other benefits to adding interactivity. For larger datasets, interactivity can allow the user to explore and discover the full dataset, going beyond the insights presented in a narrative. Aisch [2] adds that interactivity also provides transparency: since all of the data is accessible to audience, viewers can be more certain that a visualization is not depicting partial results or hiding important aspects of the data. May help increase trust in the data and the source.

## 7.2 The effect of interactivity and narrative on engagement

Through our exploratory questions, we wanted to see if our results were consistent with the results of Boy et al. [10]. They found that the total duration of time spent on the visualization (uptime) in narrative and non-narrative conditions was comparable. However, users in a visualization with an introductory narrative component with a defined narrative path were less inclined to spend time in the section of the visualization which allowed free exploration compared to users who were presented visualizations with no introductory narrative component. This introductory narrative component reduced the number of what they called meaningful interactions with the data, as well. Thus, Boy et al. conclude that the introductory narrative results in reduced data exploration, and by proxy, perhaps less insight.

Similar to their results, we did not find large differences in uptime due to the presence of narrative or interactivity. However, we offer an alternative explanation—perhaps the comparable total uptime indicates that users are only willing to spend a certain amount of time on the visualization, and this limit is not strongly affected by the design strategy employed. In the conditions without a narrative, users spent all of that time in the explore section. However, in the conditions with a narrative, users' time gets divided between the narrative section and the explore section. Thus, they necessarily spend less time in the explore section.

However, there can be two alternative explanations for the finding by Boy et al. [10]. (1) The authors tried to ensure that the narrative component did not reveal all the insights about the data, and motivated the user to explore the data by furnishing the users with questions about the data. However, it is still possible that users gained (or perceived they gained) sufficient insights about the data from the narrative. This may have led to a lower number of operations in the explore section. (2) As the user is stepping through the narrative, from each step to the next, the visual representation of the data changes. This is akin to multiple semantic operations being performed on the visualization. As a result, each *narrate* interaction may correspond to more than one interaction in the explore section—but these actions are being performed automatically to advance the narrative.

This raises another challenge in measuring the number of interactions performed by the user: one single "interaction" in one visualization may not be equivalent to a single "interaction" in another visualization (in terms of level of engagement or potential insight gained). Indeed, due to this challenge, we opted not to measure the effect of *narrative* and *interactivity* on the number of interactions in our study. In our opinion, factors such as chart types, dimensionality of the data, and design decisions regarding the type of interactions can determine the number of possible interactions for a visualization, and make comparisons across different visualization designs challenging.

## 7.3 Challenges to using engagement to measure insight

We wanted to see if engagement metrics are correlated with insight. We find that there may exist a weak linear relationship between the time spent on the visualization and the number of correctly answered questions (Figure 6A). As one would expect, spending more time with the visualization appears to increase recall and comprehension of the data. But, we also see that our participants' uptimes are comparable across the different conditions (Figure 5). These exploratory findings provide further support to our confirmatory results. Participants in the narrative conditions are more likely to answer questions correctly not because they are spending more time on the visualization, but rather because narratives make recall and comprehension of the data easier.

Figure 6B seems to indicate that there may not be any correlation between the number of interactions with the visualization and the number of correctly answered questions. Given this result and the limitations which exist in measuring interactivity across different chart types, we raise the question of whether engagement—as measured by number of interactions—is indeed a good metric to assess insight. If insight is the goal, it is simpler (and less noisy) to measure it more directly.

It is important to acknowledge that engagement may not always be a proxy for insight, it may be an end in itself. For example, it may be interesting to understand what types of designs encourage users to continually engage with a system over time. In such cases, the focus on understanding effects on engagement as an end in itself—and not as a proxy for insight—should be made clear.

## 7.4 Limitations & Future Work

The question we tested participants on naturally do not encompass all the information that can be gathered from each visualization. Readers may have have come up with their own questions from the data and generated insights to answer these questions. Our study design is unable to capture such information. A more qualitative approach might be effective in capturing information about the auxiliary insights that people may have gathered.

An alternate approach would be to use structured learning objectives for each visualization. These could be based on the designers' intended takeaways for the audience. Such learning objectives need not be limited to recall and comprehension, which comprise the lowest level of learning according to the revised Bloom's taxonomy [3]. This can also allow the learning objectives to be specific to the communicative visualization.

Finally, the participants in our study—Mechanical Turk workers—may not necessarily be interested in the topics and the data presented to them in the visualizations, unlike most readers of news-related data visualizations on the internet. We can expect most real world users to be viewing a communicative visualization due to some intrinsic motivation—for example, they might have interest in the topic of an article containing a visualization, or they might have some initial questions about the data which they hope to answer through the visualization. This can have an effect on the insights that they generate from a visualization and their engagement with it.

# 8 Conclusion

In this paper, we demonstrate the effect of narratives and interactivity on insight, measured using recall and comprehension of the data presented in the visualization. We find consistent small but positive effects on insight due the presence of narratives. On the other hand, we find that interactivity likely has little or no effect on insight. While we find that time spent on the visualization may have a weak positive correlation with recall and comprehension, the number of interactions with the visualization (a metric which is commonly used to measure engagement with visualizations) may not be correlated with recall and comprehension. Thus, we recommend caution in using these engagement metrics as a proxy for insight—if the outcome of interest is insight, measure it more directly.

In light of our results, we argue that interactivity should not be considered an essential component of communicative visualizations. Instead, if the main goal of a visualization is to provide the viewer with particular insights about the data, emphasis should be placed on design to facilitate insight generation. Put another way, if visualization design is viewed as a user-centered problem, interactivity should be added only if and when it helps the design support users' needs, not assumed *a priori* as a necessary component.

## References

[1] Gregor Aisch. Gregor aisch, information+ 2016 conference presentation. Video, Sep 2016.

[2] Gregor Aisch. In defense of interactive graphics, 2017.

[3] Lorin W Anderson, David R Krathwohl, P Airasian, K Cruikshank, R Mayer, P Pintrich, James Raths, and M Wittrock. A taxonomy for learning, teaching and assessing: A revision of bloom's taxonomy. *New York. Longman Publishing. Artz, AF, & Armour-Thomas, E.(1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. Cognition and Instruction*, 9(2):137–175, 2001.

[4] Scott Bateman, Regan L Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2573–2582. ACM, ACM, 2010.

[5] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

[6] Dominikus Baur. In defense of interactive graphics, 2017.

[7] Enrico Bertini and Moritz Stefaner. Nyt graphics and d3 with mike bostock and shan carter, May 2013.

[8] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics*, 22(1):519–528, 2016.

[9] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013.

[10] Jeremy Boy, Francoise Detienne, and Jean-Daniel Fekete. Storytelling in information visualizations: Does it engage users to explore data? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1449–1458. ACM, ACM, 2015.

[11] Paul-Christian Bürkner et al. brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1):1–28, 2016.

[12] Stuart K. Card, Jock Mackinlay, and Ben Schneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[13] Ben Casselman, Matthew Conlen, and Reuben Fischer-Baum. *Gun Deaths In America*, 2016 (accessed October 24, 2017).

[14] Remco Chang, Caroline Ziemkiewicz, Tera Marie Green, and William Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2):14–17, 2009.

[15] Kristin A Cook and James J Thomas. Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2005.

[16] Nahum Gershon and Ward Page. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37, 2001.

[17] Steve Haroz, Robert Kosara, and Steven L Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 1191–1200. ACM, ACM, 2015.

[18] Rich Harris, Nadja Popovich, and Kenton Powell. *Watch how the measles outbreak spreads when kids get vaccinated - and when they don't*, 2015 (accessed November 2, 2017).

[19] Jessica Hullman and Nick Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics*, 17(12):2231–2240, 2011.

[20] Matthew Kay, Gregory L Nelson, and Eric B Hekler. Researcher-centered design of statistics: Why bayesian statistics better fit the culture and incentives of hci. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4521–4532. ACM, 2016.

[21] Sheryl A Kluberg, Denise P McGinnis, Yulin Hswen, Maimuna S Majumder, Mauricio Santillana, and John S Brownstein. County-level assessment of united states kindergarten vaccination rates for measles mumps rubella (mmr) for the 2014–2015 school year. *Vaccine*, 35(47):6444–6450, 2017.

[22] Robert Kosara and Jock Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.

[23] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. More than telling a story: Transforming data into visually shared stories. *IEEE computer graphics and applications*, 35(5):84–90, 2015.

[24] Manuel Lima. Manuel lima, information visualization manifesto, 2009.

[25] Narges Mahyar, Sung-Hee Kim, and Bum Chul Kwon. Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life*, volume 3, page 2, 2015.

[26] S McKenna, N Henry Riche, B Lee, J Boy, and M Meyer. Visual narrative flow: Exploring factors shaping data visualization story reading experiences. In *Computer Graphics Forum*, volume 36, pages 377–387. Wiley Online Library, Wiley Online Library, 2017.

[27] Laurie Meisler. *Americans Die Younger Despite Spending the Most on Health Care*, 2017 (accessed January 12, 2018).

[28] Randall Munrow. *Earth Temperature Timeline*, 2016 (accessed Octover 14, 2017).

[29] Tamara Munzner. *Visualization analysis and design*, chapter 5, pages 85–107. CRC press, 2014.

[30] Chris North. Toward measuring visualization insight. *IEEE computer graphics and applications*, 26(3):6–9, 2006.

[31] Eric Roston, Blacki Migliozzi, Jan Dash, Yan Zhang, Blacki Migliozzi, Adam Pearce, and Mira Rojanasakul. *Bloomberg Carbon Clock: Measuring Carbon Dioxide that Causes Global Warming*, 2015 (accessed October 24, 2017).

[32] Purvi Saraiya, Chris North, and Karen Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE transactions on visualization and computer graphics*, 11(4):443–456, 2005.

[33] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16(6):1139–1148, 2010.

[34] Wita Wojtkowski and W Gregory Wojtkowski. Storytelling: its role in information visualization. In *European Systems Science Congress*, volume 5. Citeseer, Citeseer, 2002.

[35] Jo Wood, Petra Isenberg, Tobias Isenberg, Jason Dykes, Nadia Boukhelifa, and Aidan Slingsby. Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2749–2758, 2012.

[36] Ji Soo Yi, Youn ah Kang, and John Stasko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics*, 13(6):1224–1231, 2007.

[37] Ji Soo Yi, Youn ah Kang, John T Stasko, Julie A Jacko, et al. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization & Computer Graphics*, (6), 2007.

[38] Caroline Ziemkiewicz and Robert Kosara. The shaping of information by visual metaphors. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 2008.